EXERCISE 4: MULTILEVEL MODELS FOR COMPETING RISKS, USING MLwiN

In this exercise, we will be using *MLwiN* to fit multilevel discrete-time event history models for competing risks. At the same time we will allow for unobserved heterogeneity by including individual-level random effects.

We will model the outcomes of cohabitation: separation and marriage (to the same partner). In this exercise, an episode is defined as a continuous period of cohabitation. A new episode begins if a respondent starts to cohabit with a new partner.

1. Data

The data consist of those women from the subsample of 1000 NCDS respondents selected for Exercise 3 who have cohabited at some time between the ages of 16 and 42. There are 495 such women, who contribute 675 episodes of cohabitation.

The data are already in discrete-time format, with one record per six-month time interval spent in a partnership. The 495 women contribute a total of 4823 six-monthly observations. The ascii dataset **ex4.dat** contains the following variables:

WOMAN	Woman identifier
EVENT	Event indicator (0=still cohabiting, 1=separation, 2=marriage)
ATRISK	Number of months of 6-month interval for which woman exposed to risk of separation
DUR	Partnership duration, time-varying (in 6-month intervals)
PRMAR	Previously married (1=yes, 0=no)
PRCOH	Previously cohabited (1=yes, 0=no)
AGESTART	Age at start of partnership (1=less than 20, 2=20-24, 3=25-29, 4=30+)
EDUC	Number of years of post-16 education, time-varying (1=none, 2=1-2 years, 3=3-5 years, 4=6+ years)
NKIDY	Presence of preschool children, time-varying (0=none, 1=1+)
NKIDO	Presence of older children, time-varying (0=none, 1=1+)

2. Restructuring Data Prior to Fitting a Multinomial Model in *MLwiN*

To fit a multinomial model in *MLwiN*, the data must be expanded to obtain a set of binary responses for each multinomial response. Therefore, for each time interval, we have a multivariate response. This reconstruction is carried out automatically when a multinomial model is specified in *MLwiN*. The multinomial response $y_{ij}(t)$ for each time interval t is

converted to two binary responses $y_{ij}^{(1)}(t)$ and $y_{ij}^{(2)}(t)$, which respectively indicate whether separation or marriage has occurred, i.e. $y_{ij}^{(r)}(t) = 1$ if $y_{ij}(t) = r$, and $y_{ij}^{(r)}(t) = 0$ otherwise (r = 1, 2). For each time interval, the two binary responses are stacked. To illustrate this data reconstruction, consider the records for the first two episodes in our dataset:

Woman	Dur	Event	Atrisk
1	1	0	6
1	2	0	6
1	3	2	6
4	1	1	6

These episodes are contributed by two different women. The first woman married in the third 6-month interval (in the 18th month) while the second separated during the first interval (in the sixth month). Their data will restructured as follows:

Woman_long	Resp_ind	Resp	Atrisk_long	Cons.sep	Cons.mar	Dur.sep	Dur.mar
1	1	0	6	1	0	1	0
1	2	0	6	0	1	0	1
1	1	0	6	1	0	2	0
1	2	0	6	0	1	0	2
1	1	0	6	1	0	3	0
1	2	1	6	0	1	0	3
4	1	1	6	1	0	1	0
4	2	0	6	0	1	0	1

Several new variables have been created. **Resp_ind** indexes the binary responses (1 for the separation indicator, and 2 for the marriage indicator). When defining the hierarchical structure, **resp_ind** will become the level 1 identifier. The 'real' level 1 (time unit) is shifted up to level 2, with woman at level 3. **woman_long** and **atrisk_long** are 'long' versions of **woman** and **atrisk** with each value repeated twice. **Cons.sep**=1 when **resp_ind**=1 and 0 when **resp_ind**=2, and **cons.mar**=cons.sep-1. Note that the **.sep** and **.mar** suffixes come from labels attached to categories 1 and 2 of the response variable event before it was declared as a multinomial response. **Dur.sep** is **dur_long**×**cons.sep**, where **dur_long** is a 'long' version of **dur**. The **cons.*** and **dur.*** variables are created when **cons** and **dur** are added to the model as explanatory variables. The coefficients of **cons.sep** and **cons.mar** will be the intercepts in the contrasts for separation and marriage versus remaining in cohabitation, while the coefficients of **dur.sep** and **dur.mar** will be the duration effects on each contrast.

3. A Macro for Data Preparation and Setting up a Simple Model

The file **ex4_macro.txt** is a *MLwiN* macro. This macro contains syntax to read the ascii file into *MLwiN*, to create new variables, to specify a multinomial response model (after which the data is automatically restructured into a multivariate form), and to set up a simple competing model. The contents of the macro file are given at the end of this handout. All commands were previously used in Exercise 3.

We specify a simple competing risks model with duration effects, represented by a quadratic function, and age effects.

To open and run the macro in *MLwiN*:

From the **File** menu, select **Open Macro**

Locate the file **ex4_macro.txt**

Click on the **Execute** tab to run the macro

4. Fitting a Simple Competing Risks Model

To run the single-level model which ignores unobserved heterogeneity:

Open the **Equations** window (**Model → Equations**)

Click **Nonlinear** and select **Use Defaults**

Click **Start**

Click on **Estimates** to see the estimates

You should get the following results:
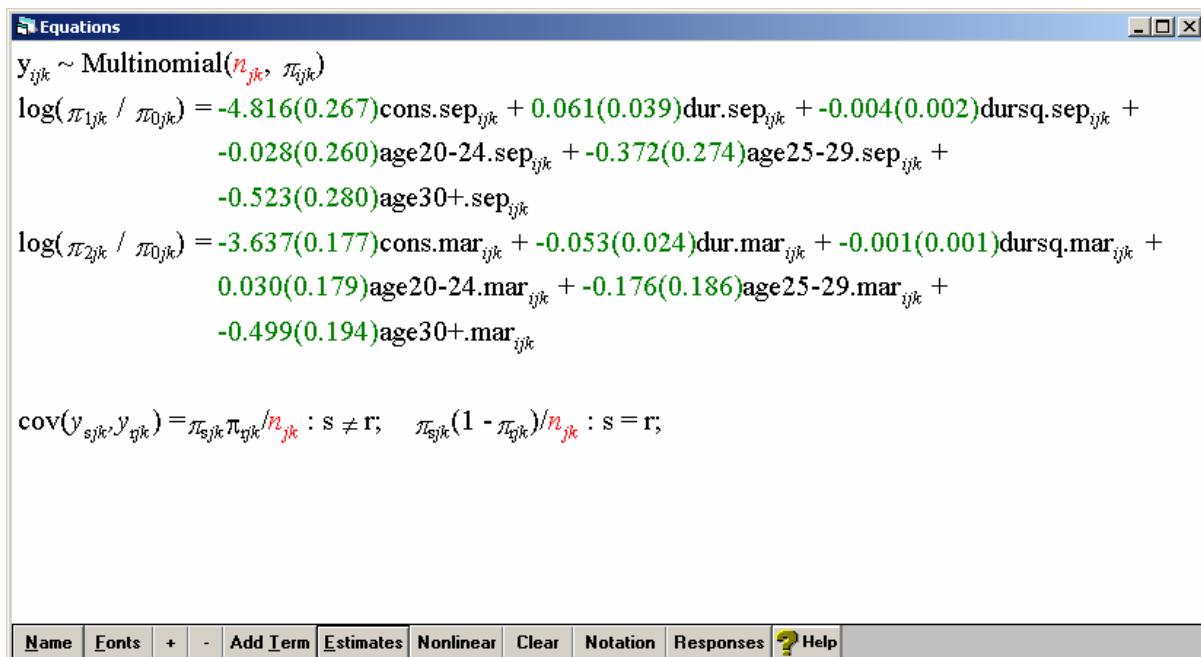
Notice that there are two equations. The first equation, which includes variables with the suffix **.sep**, is the contrast between separation and staying cohabiting. The second equation, with suffices **.mar**, is the contrast between marriage and staying cohabiting.

Is there an effect of age on the hazard of separation and/or the hazard of marriage? Carry out joint Wald tests for age, separately for each type of event.

5. Allowing for Unobserved Heterogeneity

To add random effects for each state, we need to allow the coefficients of the variables **cons.sep** and **cons.mar** to vary randomly across women.

> In the **Equations** window click on **cons.sep** (or its coefficient) and check **k(woman_long)**, then **Done**
>
> Repeat for **cons.mar**

The **Equations** window should look like this (you will need to click **Estimates** first):
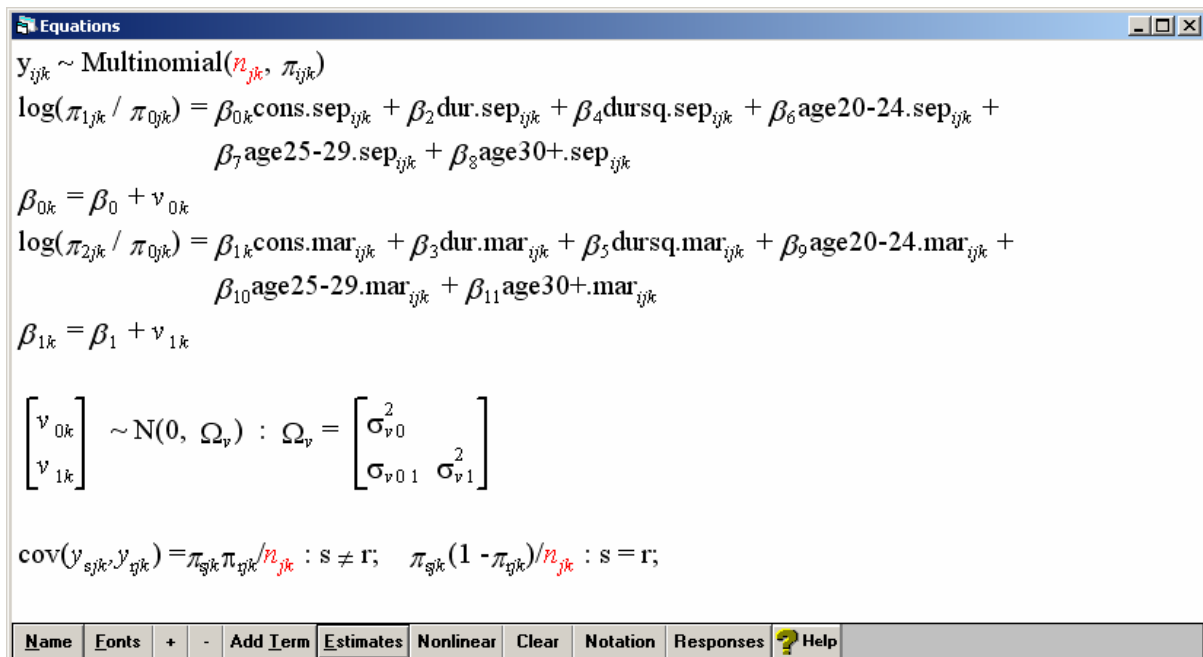
The between-woman variance in the log of the hazard of separation vs. the hazard of staying cohabiting is σ_{v0}^2. The between-woman variance for the contrast between marriage and staying cohabiting is σ_{v1}^2. The covariance between the cause-specific random effects is σ_{v01}. If women with a high propensity to separate from cohabitation have a low propensity to marry a co-resident partner, we would expect a negative covariance estimate

We will start by estimating the model using quasi-likelihood procedures, but as these are unreliable for small clusters we will switch to MCMC methods after convergence.

> Click **More** to fit the random effects model

You will get the following estimates:

Notice the zero estimate for the between-woman variance associated with marriage. Before switching to MCMC estimation, we will need to change this to a non-zero value as MCMC requires a positive definite covariance matrix for starting values. We will change this value to 0.5 as follows:

From the **Data Manipulation** menu, select **View or edit data**

Click **View** and select column **C1096**, where the current random parameter estimates are stored. The estimates are stored in lower triangular form, i.e. $(\sigma_{v0}^2, \sigma_{v01}, \sigma_{v1}^2)$.

Edit the **third row** to **0.5**, as shown below

The value for this parameter should be updated in the **Equations** window.

Now click **Estimation Control** then **MCMC**

We will run MCMC with the default settings. i.e. a burn-in of 500 and a chain length of 5000

Click **Done** to activate the MCMC settings, then **Start**

MCMC will take a few minutes to run. In the meantime you may wish to start looking at the exercises below.

You should get the following results:

Again we need to check whether convergence has been achieved before interpreting the results. If we look at the trajectories, we see that mixing is extremely poor. Below are the results from a burn-in of 5000 and a chain length of 50,000.

Note that the estimated covariance between the random effects is close to zero with a large standard error. We therefore conclude that there is little evidence of correlation between the unobserved woman-level characteristics affecting separation and those affecting the hazard of a transition to marriage.

6. Exercises

Because of the length of time required to estimate models using MCMC, we will ignore unobserved heterogeneity for now. In practice, it is advisable to carry out preliminary analysis using single-level models, fitting only selected models with random effects. Alternatively, the estimates obtained using quasi-likelihood methods are useful as an approximation.

To remove the random effects:

Click **Estimation Control** then **IGLS/RIGLS**, then **Done**

Click on **cons.sep** and deselect **k(woman_long)**

Repeat for **cons.mar**. The woman-level random effect covariance matrix should disappear

Click **Start** to refit the single-level model

- Modify **ex4_macro.txt** to add in other covariates to the model. Remember to declare variables with more than two categories as categorical.

- Carry out significance tests for these covariates. Which factors affect the hazard of separation, and which the hazard of marriage?

1	NOTE: Exercise 4 - macro to set up competing risks model in MLwiN
2	dinp c1-c10
3	ex4.dat
4	name c1 'woman' c2 'event' c3 'atrisk' c4 'dur' c5 'prmar' c6 'prcoh'
5	name c7 'agestart' c8 'educ' c9 'nkidy' c10 'nkido'
6	NOTE: Calculate duration-squared variable
7	calc c11='dur'*'dur'
8	name c11 'dursq'
9	NOTE: Create level 1 ID (coded 1,2, . . ., 4823)
10	code 4823 1 1 c12
11	name c12 'lev1id'
12	NOTE: Create vector of ones - needed for estimation of intercept terms
13	put 4823 1 c13
14	name c13 'cons'

15	NOTE: Declare EVENT as categorical then as a multinomial response, with category 0 the reference
16	NOTE: 2 new variables are created, the stacked binary responses for categories 1 and 2
17	NOTE: and a response index
18	catn 1 'event' 0 'none' 1 'sep' 2 'mar'
19	mnom 0 'event' c14 c15 0
20	name c14 'resp' c15 'resp_ind'
21	NOTE: Create 'long' version of ATRISK. (Values of other variables will be repeated automatically
22	NOTE: when added to the multinomial model.)
23	repe 2 'atrisk' c17
24	name c17 'atrisk_long'
25	NOTE: Divide by ATRISK to obtain probability of each type of event within 6-month interval and
26	NOTE: declare as response
27	calc c18='resp'/'atrisk_long'
28	name c18 'y'
29	resp 'y'
30	NOTE: Declare ATRISK as the denominator
31	calc c19='atrisk_long'
32	name c19 'denom'
33	NOTE: Declare multilevel structure. Level 1 used to define multivariate structure.
34	iden 1 'resp_ind' 2 'lev1id' 3 'woman'
35	NOTE: Declare age variables as categorical and name categories
36	catn 1 'agestart' 1 'age<20' 2 'age20-24' 3 'age25-29' 4 'age30+'
37	NOTE: Add covariates
38	addt 'cons'
39	addt 'dur'
40	addt 'dursq'
41	addt 'agestart'